Vehicular Networks on Two Madrid Highways

Marco Gramaglia, Oscar Trullols-Cruces, Diala Naboulsi, Marco Fiore, Maria Calderon

RhôneAlpes Région

informatics mathematics

Vehicular Networks on Two Madrid Highways

Marco Gramaglia, Oscar Trullols-Cruces, Diala Naboulsi, Marco Fiore, Maria Calderon

TECHNICAL REPORT

N°

March 2014

Project-Teams

ISSN 0249-0803 ISRN INRIA/RT---FR+ENG

Vehicular Networks on Two Madrid Highways

Marco Gramaglia*, Oscar Trullols-Cruces†, Diala Naboulsi‡,
Marco Fiore§, Maria Calderon ¶

Project-Teams

Technical Report n° — March 2014 — 18 pages

Abstract: There is a growing need for vehicular mobility datasets that can be employed in the simulative evaluation of protocols and architectures designed for upcoming vehicular networks. Such datasets should be realistic, publicly available, and heterogeneous, i.e., they should capture varied traffic conditions. In this paper, we contribute to the ongoing effort to define such mobility scenarios by introducing a novel set of traces for vehicular network simulation. Our traces are derived from high-resolution real-world traffic counts, and describe the road traffic on two highways around Madrid, Spain, at several hours of different working days. We provide a thorough discussion of the real-world data underlying our study, and of the synthetic trace generation process. Finally, we assess the potential impact of our dataset on networking studies, by characterizing the connectivity of vehicular networks built on the different traces. Our results underscore the dramatic impact that relatively small communication range variations have on the network. Also, they unveil previously unknown temporal dynamics of the topology of highway vehicular networks, and identify their causes.

Key-words: vehicle-to-vehicle communications, vehicular networks, mobility modeling, connectivity analysis

Funding for Diala Naboulsi was provided by a grant from the Rhône-Alpes Region.

* ISMB, Torino, Italy
† Universitat Politècnica de Catalunya, Barcelona, Spain
‡ INSA Lyon / Inria, Lyon, France
§ CNR-IEIIT / Inria, Torino, Italy
¶ Universidad Carlos III, Madrid, Spain

RESEARCH CENTRE
GRENOBLE – RHÔNE-ALPES

Inovallée
655 avenue de l'Europe Montbonnot
38334 Saint Ismier Cedex

Réseaux Véhiculaires sur des Autoroutes de Madrid

Résumé : Il y a un besoin croissant de jeux de données de mobilité pour l'évaluation d'architectures et de protocoles réseaux conçus pour les environnements véhiculaires. Tels jeux de données doivent être réalistes, capables de refléter les variétés des conditions routières et accessibles à la communauté scientifique. Dans ce travail, nous introduisons un ensemble de traces de mobilité pour la simulation des réseaux véhiculaires qui respectent les pré-requis cités. Ces traces sont dérivées en se basant sur des mesures à haute résolution des flux de trafic réels sur deux autoroutes autour de la ville de Madrid, en Espagne. Nous évaluons leur intéret en caractérisant la connectivité des réseaux véhiculaires pour les différentes traces. Nos résultats soulignent l'impact important des variations de la portée radio entre les véhicules. En outre, ils dévoilent les dynamiques temporelles de la topologie des réseaux véhiculaires en environnement autouroutier, et en identifient les causes.

Mots-clés : communications entre vehicules, réseaux vehiculaires, modélisation de mobilité, analyse de connectivité

1 Introduction

Vehicular communications are regarded as a key enabling technology within upcoming Intelligent Transportation Systems (ITS), which are in turn expected to significantly improve road safety and traffic management. Indeed, communication-enabled vehicles will be able to exchange data in a direct way, via the so-called vehicle-to-vehicle (V2V) communication, so as to build spontaneous, distributed, self-organizing vehicular networks. The latter are anticipated to complement the existing cellular network, by providing short-range, low-latency, inexpensive transfers that well integrate today's pervasive but strongly centralized radio access network architecture. The same features also make vehicular networks especially suitable to implement ITS services such as cooperative awareness, collision avoidance, and danger warning.

The moment when V2V technologies and the associated services will hit the market is in fact approaching. Standards especially designed for V2V communication, such as IEEE 802.11p, IEEE 1609 OSI CALM-M5 and ETSI ITS are being finalized, and the automotive industry is actively participating in the first field tests, such as those carried out via the sim$^{\text{TD}}$ platform in Frankfurt, Germany, or through the Ann Arbor Safety Pilot in Michigan, USA.

However, experimental trials remain an exception, due to their costs and complexity. The vast majority of applications, protocols and architectures for upcoming vehicular networks is still evaluated via simulative studies. Within such a context, the level of realism of the simulation is a paramount aspect to account for, and the way the mobility of individual vehicles is represented is often the single feature that introduces the largest bias in the results [1].

During the past decade, there have been efforts aiming at gathering real-world road traffic data [2–4], developing tools for the simulation of vehicular movement [5, 6], and generating synthetic mobility datasets [7–9]. This notwithstanding, there is still a noticeable scarcity of road traffic datasets featuring the level of realism and of spatiotemporal granularity required for network simulation [10]. The result of the lack of a set of realistic, publicly shared, heterogeneous scenarios is that simulations of vehicular networks are often unreliable and non-reproducible [11].

In this paper, we contribute to the endeavor of enlarging the set of realistic datasets of vehicular mobility that are freely available to the research community. To that end, we present 16 original traces describing the road traffic on two highways around Madrid, Spain, at different times of multiple weekdays. The traces are derived from high-detail real-world traffic counts, and describe unidirectional free-flow traffic in quasi-stationary conditions on a road spanning over 10 km. More precisely, our contributions are as follows:

- We collect real-world traffic count data that feature an unprecedented resolution, and study the road traffic dynamics they capture. Our analysis pinpoints the heterogeneity of the empirical dataset, and confirms that widely adopted assumptions on exponential inter-arrivals of vehicles do not necessarily hold in real scenarios.

- We present a process to generate synthetic mobility traces from road traffic counts, by leveraging microscopic models and properties of the real-world data. Our analysis is relevant to the calibration of car-following and lane-changing models, and results in original, publicly available datasets of highway mobility for network simulation.

- We provide a characterization of the vehicular network connectivity in the synthetic traces, investigating the impact of time (i.e., hour of the day, day of the week), highway diversity, and vehicular communication range. Our results highlight that: (i) the level of global vehicular connectivity can be primarily ascribed to the V2V communication range; (ii) for short communication ranges, the vehicular connectivity is driven by a mixture of slow and fast temporal dynamics that are regulated by the road traffic density and by V2V contact durations, respectively.

The paper is organized as follows. After a discussion of related works, in Sec. 2, we present and analyze the real-world traffic count data we use in our study, in Sec. 3. Then, Sec. 4 describes the generation process of the synthetic vehicular mobility traces, whose vehicular network connectivity is characterized in Sec. 5. Finally, Sec. 6 concludes the paper.

2 Related Work

The impact of realistic mobility models on the simulation of communication protocols for vehicular networks has been emphasized in many works [1,8,11,12]. This fact has pushed the research community to seek for an ever-increasing realism in road traffic traces used to feed network simulators.

A first approach is to use real-world mobility traces that have been recorded by directly logging the position of vehicles during their movements. Unfortunately, the traces of this kind that are currently available yield a partial description of the road traffic, as they only report the positions of, e.g., buses [2, 3] or taxis [4]. Also, they are typically affected by limited temporal accuracy, with positions logged with periodicities in the order of tens of seconds [10].

Therefore, most works have focused instead on the generation of synthetic vehicular traces. In particular, a number of studies have considered urban mobility scenarios, created by feeding real-world road topologies of different cities to microscopic traffic simulators such as SUMO [5] or VanetMobiSim [6]. In order to characterize the number, origin, destination and time of trips, these works usually made use of macroscopic data (i.e., origin-destination matrices) collected from user surveys [7, 8, 13] or from roadside detectors such as induction loops, cameras and infrared counters [9]. However, the dynamics of traffic in urban regions – characterized by vehicles traveling at low or medium speed, and frequent intersections regulated by traffic lights or roundabouts – are not comparable to those on highways – featuring instead high speeds and frequent overtakings. Moreover, none of the aforementioned works considers the problem of fine-tuning the microscopic mobility model, as we do in this study.

Closer to our approach, the study in [12] uses two empirical road traffic datasets to generate synthetic highway mobility traces that can be used in network simulation. The two 24-hour datasets were collected using detectors underneath the road: the first dataset is from the I-80 highway nearby Berkeley, CA, USA, and contains information from dual-loop detectors (that log data on individual vehicles passing through); the second dataset is from the Gardiner Expressway in Toronto, Canada, and includes records from metal detectors (that measure aggregated characteristics on each lane every 20 seconds). In order to generate the synthetic traces, inter-arrivals are assumed to follow exponential distributions and vehicle speeds are modeled as Gaussian distributions: the real-world data is then employed to calculate the parameters of such distributions for different hours of a day. Then, a stochastic vehicle mobility generator is developed according these models. Our study improves that in [12] from several viewpoints: (i) we consider more detailed and heterogeneous traffic count data; (ii) we use the empirical traffic counts as they are, and not to parameterize distributions; (iii) we employ validated microscopic car-following and lane-changing models instead of resorting to stochastic modeling; (iv) the synthetic mobility traces we generate are publicly available, which is not the case for those of I-80 and Gardiner Expressway.

Highway scenarios are considered also in [14], where empirical data from the Freeway Performance Measurement System (PeMS) [1] is fed to the SUMO microscopic mobility simulator to generate synthetic traces. In particular, the real-world data, collected from road sensors in Alameda County, CA, USA, is used to determine the assignment of vehicular traffic flows and

[1] `http://pems.dot.ca.gov/`

(a) (b) (c) (d)

Figure 1: (a): Geographical location of the two highways near Madrid,Spain. (b): Close-by view of measurement point on M40. (c): Close-by view of measurement point on A6. (d): Close-by view of the induction loop on M40.

the average speed values over the road. However, the empirical traffic count dataset used in [14] is affected by a coarse time granularity (i.e., flow and speed sampling rates between 30 seconds and 5 minutes), which prevents using it to model the precise arrival time of each vehicle, as we do instead.

Also relevant to our work are vehicular network connectivity studies. There, several research teams have focused on urban areas [4, 15–17]. Although interesting, their findings are hardly applicable to the highway scenarios we are interested in, due again to their significant differences with respect to urban environments. As far as highway traffic is concerned, a number of works dealt with the analytical characterization of vehicular connectivity in unidimensional scenarios [18–20], which however implies adopting simplifying assumptions in order to make the problem mathematically tractable.

Fewer papers rely instead on realistic mobility traces. Specifically, in [21] synthetic data generated via a microscopic simulator is used to confirm the intuition that increasing the traffic density or the number of lanes improves connectivity. Nevertheless, the study employs a simple highway traffic model with deterministic vehicle inter-arrivals, while our analysis relies on much more realistic mobility traces built from real-world data. More realistic mobility is accounted for in [22], which uses (publicly unavailable) reality-audited highway movements from the FleetNet project, and in [14], which employs the aforementioned Alameda County trace. In both cases, however, the connectivity assessment is limited to statistical analyses of node degree and link duration distributions. In this paper, we provide a more thorough characterization of the instantaneous vehicular network connectivity, including original quantitative results on the correlation between the radio communication range, the traffic density, and the number and size of network components.

3 Real-world traffic count data

The empirical data used throughout this work comes from measurement locations in the region of Madrid, Spain. The data, kindly provided to us by the Spanish office for the traffic management (Dirección General de Tráfico, DGT), detail road traffic conditions on the following two arterial highways.

- **M40**: the M40 is part of the middle layer of the Madrid city beltway system, which also comprises M-30 (innermost) and M-50 (outermost). This beltway, which has an average distance of 10.7 km from the city center, traverses both the most peripheral areas of the municipality and surrounding cities. The measurement point is placed at the 12.7-km milepost, where the M40 traverses the suburb of San Blas and the town of Coslada. The

measures cover the internal carriageway (southbound) that includes 3 lanes with a speed limit of 100 km/h. The M40 is indicated as A in Fig. 1(a), while a photographic view of the exact measurement location is provided in Fig. 1(b).

- **A6**: the Autovía A6 is a motorway that connects the city of A Coruña to the city of Madrid. This road enters into the urban area from the northwest and collects the traffic demand of the conurbation that was built along it. The data collection point is placed around the 11-km milepost (Madrid direction), where the A6 features three lanes with a speed limit of 120 km/h. We remark, however, that this location is right after a popular entrance ramp that joins the rightmost lane and significantly slows down the road traffic. The highway maps to B in Fig. 1(a). A photographic view of the collection point is in Fig. 1(c).

3.1 Collecting high-resolution traffic counts

The sensors that DGT deployed on M40 and A6 are induction loops. They consist of loops of wires buried under the concrete layer that create a magnetic field. When a vehicle passes on the vertical axis of the loop, it creates a variation in the magnetic field, which is considered as a new transit. If two loops are placed close to each other, also the vehicular speed (and possibly other metrics, such as the vehicle length) is obtained. A picture of the buried induction loops used to collect measurements on the M40 is provided in Fig. 1(d).

Usually, these devices are programmed to supply coarse-grained data, as public transportation authorities are generally interested in aggregate measures [2] (e.g., the number of vehicles transiting on a road, their average speed, or the percentage of heavy vehicles) so as to detect exceptional alterations in the road capacity [12,14]. The loops installed by DGT were configured to supply data averaged over 60-second intervals, but their configuration was changed specifically for our study: the real-world traffic counts we collected provide fine-grained time, speed and lane information on each single transiting vehicle. To the best of our knowledge, this is the most precise traffic count dataset recorded to date, and represents an ideal input to our microscopic simulation of highway traffic.

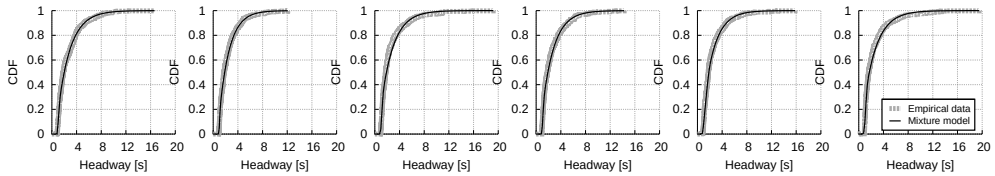
The timing of the data collection is a very important variable to account for, since vehicular traffic presents, as many other human activities, important temporal variability: for instance, rush hours yield diverse traffic conditions than off-peak hours, especially on main arterial roads like those we consider. The dataset analyzed in this work was collected on multiple days of May 2010, namely Friday the 7th, Monday the 10th, Tuesday the 11th, and Wednesday the 12th. For each measurement location and each day, two traces are available: one capturing the traffic peak (from 8:30 a.m. to 9:00 a.m.), and one during off-peak hours (from 11:30 a.m. to 12:00 p.m.). As a result, our empirical dataset provides a heterogeneous view of traffic conditions, allowing us to generate synthetic mobility traces that are representative of different highway congestion levels.

3.2 Understanding the data

Each entry of the dataset refers to one individual vehicle transiting over the induction loops, and includes the following:

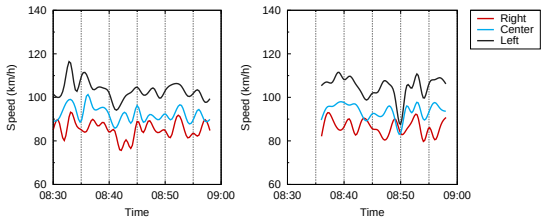
- **Timestamp**: the time at which the vehicle passage was recorded by the induction loop. The precision of the time reference is 100 milliseconds.

[2] As an example, Dirección General de Tráfico provides an elaboration of the traffic data via the Infocar web service at `http://infocar.dgt.es`, including historical data visualization for some of the observation points.



(a) M40, right lane (b) M40, center lane (c) M40, left lane (d) A6, right lane (e) A6, center lane (f) A6, left lane

Figure 2: Inter-arrival times CDF measured on May 12, 2010. Each plot refers to a different lane of M40 at 8:30 a.m. (a, b, c) and A6 at 11:30 a.m. (d, e, f). Solid black lines show the mixture model for each distribution. The fitting is good under any combination of highway, day and hour for each lane.



(a) Tue 11, 8:30 a.m. (b) Mon 10, 8:30 a.m.

Figure 3: Per-lane speed versus time in two dataset samples. The plots are representative of free flow traffic (a) and of a mild perturbation (b) on M40.

- **Speed**: the vehicle speed, in km/h.
- **Lane**: the lane on which the vehicle transited. Both M40 and A6 feature three lanes at the measurement points. They are referred to as the right, center and left lanes (in the heading direction) in the remainder of the paper.

In [23], we performed an extensive analysis of vehicle inter-arrival times (IAT) in the traffic count dataset. We found that, although many works assume exponential distribution or even uniform distribution to model the time headway between subsequent vehicles, a mixture gaussian-exponential model yields a much better approximation of the empirical data. The mixture model also provides valuable information on drivers' behavior: the gaussian part of the distribution captures *bursty* arrivals of vehicles that travel close to each other at similar speeds, whereas the exponential part models *isolated* vehicles whose movement is less constrained by that of other cars. Fig. 2 shows the excellent match between the mixture model and the experimental data on multiple combinations of highway, day and hour for each lane.

An overview of two traffic count data samples is provided in Fig. 3. Each plot depicts the evolution over time of the speed on the three lanes of the highway under consideration: consistently with Spanish road regulations, the speed of the right lane is typically the lowest, while that of the left lane is normally the highest. As anticipated, the dataset used in this work presents different traffic characteristics in terms of traffic measures (i.e., density, flow and speed). However, we remark that, throughout all the dataset, the road traffic is always in a so-called *free flow* state, i.e., there is a quasi-linear relationship between the vehicular flow and density, and no significant congestion is observed [24]. Free flow traffic conditions result in neatly separated speeds on different lanes, as shown, e.g., in Fig. 3(a). Only in rare occasions, we record mild congestion situations that result in a *pinch effect*, i.e., travel speeds on different lanes are leveled down to a same value [24]. An example of such a phenomenon is observed in Fig. 3(b), at around 8:50 a.m.

4 Road traffic dataset generation

Our objective is to generate road traffic traces that are representative of unidirectional free flow highway traffic in a quasi-stationary state, i.e., such that comparable traffic conditions are present in between the in-flow and out-flow boundaries of the simulated highways. The datasets should cover a geographical distance that is sufficient for networking studies (e.g., for warning message dissemination or floating car data upload). In the following, we simulate a 10-km road segment, but we remark that our approach can easily accommodate shorter or longer highway stretches.

In order to achieve the goal above, we fed the real-world traffic count data presented in Sec. 3 to a microscopic vehicular mobility simulator. Next, we first present the state-of-art car-following and lane-changing models employed by the simulator. Then, we discuss the calibration of the microscopic models to attain quasi-stationary road traffic conditions. Finally, we outline the features of the resulting dataset.

4.1 Microscopic vehicular mobility models

The car-following and lane-changing microscopic mobility models implemented by our simulator are IDM and MOBIL, both of which have been validated by the transportation research community and are among those most commonly used in vehicular networking research.

The Intelligent Driver Model (IDM) [25] characterizes the behavior of the driver of a car i through her instantaneous acceleration $dv_i(t)/dt$, which is calculated as

$$\frac{dv_i(t)}{dt} = a\left[1-\left(\frac{v_i(t)}{v_i^{max}}\right)^4-\left(\frac{\Delta x_{des}(t)}{\Delta x_i(t)}\right)^2\right] \quad (1)$$

$$\Delta x_{des}(t) = \Delta x_{safe}+\left[v_i(t)\Delta t_{safe}-\frac{v_i(t)\Delta v_i(t)}{2\sqrt{ab}}\right] \quad (2)$$

In (1), $v_i(t)$ is the current speed of vehicle i, v_i^{max} is the maximum speed its driver would like to travel at, and $\Delta x_{des}(t)$ is the so-called *desired dynamical distance*, representing the distance that the driver should keep from her leading vehicle. The latter is computed in (2) as a function of several measures taken with respect to the car in front of vehicle i: the minimum bumper-to-bumper distance Δx_{safe}, the minimum safe time headway Δt_{safe}, and the speed difference $\Delta v_i(t)$. In both equations, a and b denote the maximum absolute acceleration and deceleration, respectively. When combined, these formulae return the instantaneous acceleration of the car, as a combination of the desired acceleration $[1-(v_i(t)/v_i^{max})^4]$ on a free road, and the braking deceleration induced by the preceding vehicle $(\Delta x_{des}(t)/\Delta x_i(t))^2$.

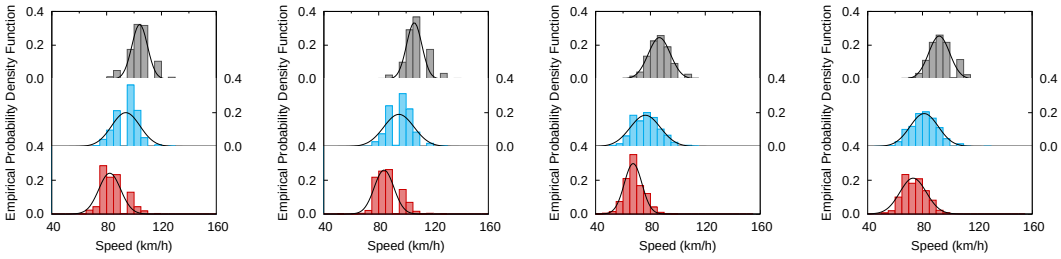
The Minimizing Overall Braking Induced by Lane-changes model (MOBIL) [26], follows a game theoretical approach, and allows the driver of a vehicle i to move to an adjacent lane if her advantage in doing so is greater than the disadvantage of the trailing car j in the new lane. The (dis)advantage is measured in terms of acceleration, which translates the lane movement condition above into the inequality

$$\left|\frac{dv_i(t)}{dt}\right|_L-\frac{dv_i(t)}{dt}+a_L \geq p\left(\frac{dv_j(t)}{dt}-\left|\frac{dv_j(t)}{dt}\right|_L\right)+ka, \quad (3)$$

where the notation $|\cdot|_L$ denotes accelerations computed as if vehicle i were traveling on the lane to its left rather than in the current one. In (3), $p \in (0,1]$ is a politeness factor that models the selfishness of the driver with respect to the new back vehicle j, ka is a hysteresis threshold (computed as a fraction k of the maximum acceleration a) that prevents lane hopping, and a_L is

Table 1: IDM and MOBIL parameter settings

Model	Parameter	Meaning	Value
IDM	a	Maximum acceleration	1 m/s^2
IDM	b	Maximum (absolute) deceleration	2.5 m/s^2
IDM	v_i^{max}	Maximum desired speed	$\sim f_V(v)$
IDM	Δx_{safe}	Minimum safety distance	1 m
IDM	Δt_{safe}	Minimum safe time headway	0.65 s
MOBIL	p	Politeness factor	0.5
MOBIL	a_L	Bias acceleration (left)	0 m/s^2
MOBIL	a_R	Bias acceleration (right)	0.2 m/s^2
MOBIL	k	Hysteresis threshold factor	0.3



(a) M40, 8:30 a.m. (b) M40, 11:30 a.m. (c) A6, 8:30 a.m. (d) A6, 11:30 a.m.

Figure 4: Empirical per-lane ingress speed distributions and fitted normal distributions for different highway and measurement hour combinations.

a bias acceleration that can be used to favor or limit movements to left. An identical formulation can be used for right-hand-side lane changes, and the respective advantages can be compared to determine the final lane movement, if any. Note that, in Spain, road traffic regulations enforce drivers to travel on the rightmost lane whenever possible: we thus expect $a_R > a_L$ and $a_R > 0$, i.e., right-hand-side movements to be favored over left or no movement, if equivalent conditions are present on all lanes.

4.2 Parameter calibration

In order to obtain quasi-stationary traffic conditions over the simulated highway segment, some calibration of the IDM and MOBIL parameters proved necessary. Specifically, for the acceleration a, deceleration b, politeness factor p and minimum safety distance Δx_{safe} we could use the default values suggested in [25, 26]. The other parameters had instead to be adapted to the specificities of each road scenario, as detailed next. We remark that we employed a discrete time step of 100 ms in all mobility simulations. The final IDM and MOBIL settings are summarized in Tab. 1.

Maximum desired speed. Vehicles enter the simulated highway segment at the time and with the speed defined by the real-world traffic count data. However, it is then necessary to determine the maximum desired speed v_i^{max} of each driver, i.e., the cruise velocity that she would keep if she were alone on the highway. To that end, we exploit our observation in Sec. 3.2 that all traffic count datasets yield free flow conditions. We thus assume that the experimental distribution of real-world ingress speeds can be used as a base for inferring the actual distribution of desired speeds.

We thus computed the Probability Density Functions (PDF) of traffic count ingress speeds for each combination of road, day, and measurement hour. We found that the empirical distributions for the same highway at the same time were overlapping, and thus aggregated such data for different days. This left us with four combinations of highway and hour. For each such combination, we separated the results on a per-lane basis, since each lane yields significantly diverse speeds.

The resulting PDFs are shown in Fig. 4. We note that the empirical ingress velocities tend to follow gaussian distributions in all cases. Nonlinear least squares fittings performed via the Trust-Region algorithm confirm that intuition, with adjusted R-square values always above 0.97 and root mean square errors below 0.02. The fitted gaussian distributions are portrayed as solid lines in the plots of Fig. 4.

Despite being measured in free flow conditions, ingress speeds cannot be directly mapped to desired velocities. Indeed, the presence of surrounding vehicles implicitly limits the ingress speeds, even if by a limited extent. Therefore, when the values of v_i^{max} are directly extracted from the gaussian ingress speed distributions, vehicles tend to have an egress velocity from the simulated road segment that is on average 2.8 m/s (around 10 km/h) lower than expected. By adding such a speed offset to the mean of the theoretical PDF, we obtain, for each lane of every highway and hour combination, a distribution of maximum desired speed $f_V(v)$ that grants quasi-stationary speed traffic over the whole highway stretch. Interestingly, we highlight that the shapes of such $f_V(v)$ are similar for the two measurement hours on a same highway, but remarkably different for the two highways[3]: this matches the intuition that the desired speed of a driver is influenced by the type of road she is traveling on, rather than by daytime.

Safety time. The safe time headway Δt_{safe} was reduced from the default value of 1.5 s to 0.65 s. As a matter of fact, we observed that in the real-world traffic count data the inter-arrival time between two cars on a same lane may get much smaller than $1.5s$. Maintaining the default value of Δt_{safe} can then bias the IDM behavior in cars that just entered the scenario. Specifically, we recorded excessive brakings that caused traffic jams at the entrance, which in turn prevented new cars to enter the highway segment as commended by the real-world data. After extensive tests, we selected $\Delta t_{safe} = 0.65$ s, as that is the largest safety time for which the effect described above disappears, allowing vehicles to enter the simulated scenario according to the traffic count datasets.

Lane change bias and hysteresis threshold. The default MOBIL settings resulted in a traffic that was highly skewed towards the left lane, which became then unrealistically congested. We run comprehensive trials to find the combination of right (a_R) and left (a_L) lane change bias, and lane change hysteresis threshold factor (k) that granted quasi-stationary traffic over the different lanes. The most consistent ingress and egress per-lane properties were obtained for $a_R = 0.2$ m/s^2, $a_L = 0$ m/s^2, and $k = 0.3$. Those values also respect the intuitive observation that traffic should be skewed to the right so as to abide Spanish road regulations, and we thus chose to use them for the generation of the traces.

4.3 Final dataset

The final road traffic dataset [4] is composed of 16 different mobility traces, resulting from all possible combinations of two highways, four days and two measurement times. By default, the position of each vehicle is recorded over a 10-km road stretch at every second. Each mobility trace has a duration of 30 minutes, dictated by the measurement intervals of the real-world traffic

[3] This can be observed by looking at the gaussian shapes in Fig. 4, since the final $f_V(v)$ are just shifted versions of the same over the x axis.

[4] Available at `http://trullols.site.ac.upc.edu/downloads`.

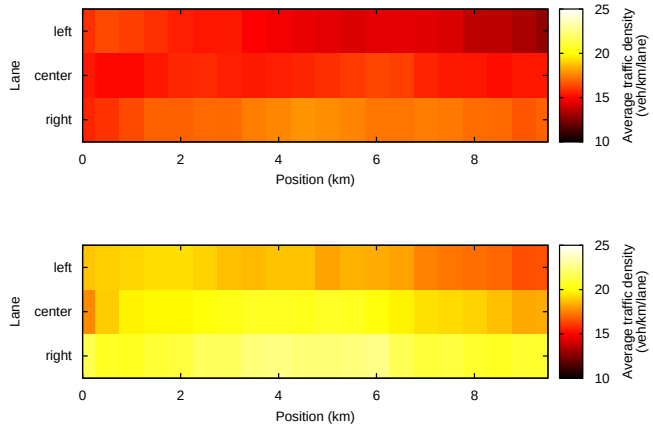


Figure 5: Heatmaps of the average vehicular density in two sample traces, on Wednesday the 12th: M40 at 11:30 a.m. (top) and A6 at 8:30 a.m. (bottom)

count data. As mentioned earlier, all traces are representative of quasi-stationary road traffic. For instance, Fig. 5 presents heatmaps of the average vehicular density recorded on two sample traces: we can remark that traffic conditions are similar all over the road stretch (along the x axis of the plots) for each lane (y axis of the plots).

Finally, Tab. 2 summarizes the main properties of the 16 traces, in terms of average values of speed, vehicular density, and headway time recorded over the 30-minute duration of each trace. As already anticipated, the speeds measured on each highway at the same time of different days tend to be similar. However, different highways and hours yield significant diversity in terms of average speeds, which range between 68 km/h (A6 at 8:00 a.m.) and 87 km/h (M40 at 11:30 a.m.). Also, we remark in all traces the correlation between the speed and the other metrics, which is consistent with well-known road traffic theory results [24]. However, M40 and A6 show non-comparable behaviors in that sense, since lower densities and longer headway times on A6 can lead to a traffic that – counter-intuitively – is much slower than that recorded on M40 in presence of higher density and shorter headway times. This is an effect induced by the different nature of the two highways, and a proof of the heterogeneous nature of the traces in our vehicular mobility dataset.

5 Connectivity analysis

In this section, we study the connectivity properties of the vehicular networks built on the mobility traces derived in the previous sections. We start by presenting the model that we apply and the metrics that we consider, followed by the results we obtained. We conclude the section by discussing vehicle-to-vehicle contact duration properties.

5.1 Network model and metrics

In our analysis, we focus on the instantaneous connectivity properties of the network. For each time instant, we represent the vehicular network as an undirected graph $G(\mathbb{V}(t), \mathbb{E}(t))$, where $\mathbb{V}(t) = \{v_i(t)\}$ is the set of vertices $v_i(t)$, representing each a vehicle i appearing in the network at time t, and $\mathbb{E}(t) = \{e_{ij}(t)\}$ is the set of edges $e_{ij}(t)$. We consider that an edge $e_{ij}(t)$ is present

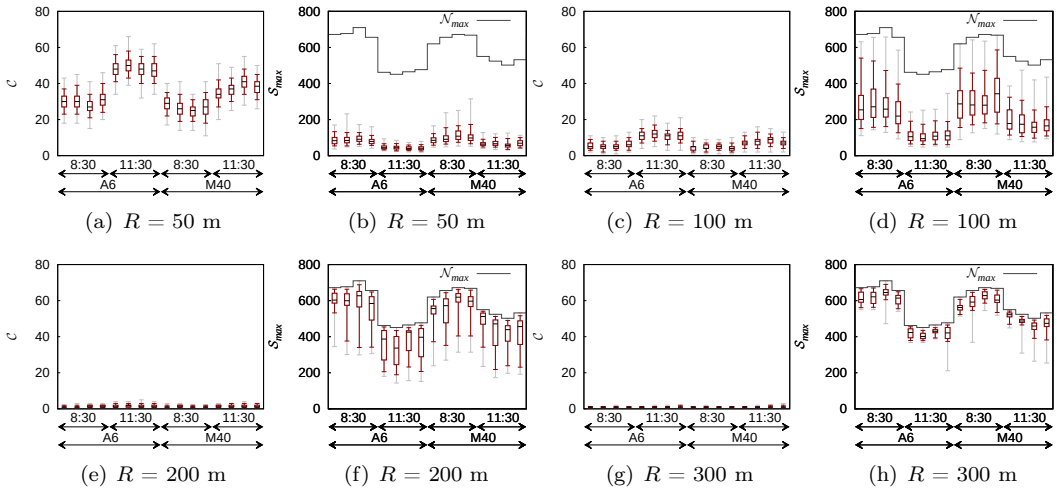
Table 2: Main average characteristics of the 16 traces

Trace	Speed (km/h)	Density (vehicles/km)	Headway (s)
M40, Fri 7, 8:30	85.66	56	0.74
M40, Mon 10, 8:30	84.85	59	0.72
M40, Tue 11, 8:30	84.33	63	0.70
M40, Wed 12, 8:30	84.98	60	0.71
M40, Fri 7, 11:30	86.88	52	0.82
M40, Mon 10, 11:30	87.49	49	0.84
M40, Tue 11, 11:30	87.52	46	0.88
M40, Wed 12, 11:30	87.71	48	0.85
A6, Fri 7, 8:30	69.09	62	0.86
A6, Mon 10, 8:30	68.73	62	0.85
A6, Tue 11, 8:30	68.24	65	0.83
A6, Wed 12, 8:30	68.65	61	0.86
A6, Fri 7, 11:30	77.26	42	1.11
A6, Mon 10, 11:30	77.01	40	1.14
A6, Tue 11, 11:30	76.99	43	1.10
A6, Wed 12, 11:30	77.19	42	1.13

between $v_i(t)$ and $v_j(t)$, if a communication link exists between vehicles i and j at time t. We adopt the unit disc model as a representation of the RF signal propagation. Thus, we assume that two vehicles can communicate directly if they are separated by a distance of at most R meters, where R is the communication range. We apply this simplistic model due to the fact that deterministic propagation models do not scale well to large mobile scenarios and imply expensive instantaneous computations. Moreover, the unit disc model is representative of the average behavior of the system obtained under stochastic propagation models. Specifically, in the following we present results for different values of R, from 50 to 300 meters. Despite standards claiming up to 1-km communication ranges, these values represent the lower and upper bounds for acceptable packet delivery ratios, according to different experimental studies on Dedicated Short Range Communication (DSRC) transfers [27, 28].

We use the graph model to define several metrics of interest to the instantaneous connectivity analysis. Namely, a component $C_k(t) = G(\mathbb{V}_k(t), \mathbb{E}_k(t))$ is a subgraph of $G(\mathbb{V}(t), \mathbb{E}(t))$, where $\mathbb{V}_k(t)$ is a subset of $\mathbb{V}(t)$ and includes a group k of vehicles that can communicate together via direct or multi-hop links at time t, and $\mathbb{E}_k(t)$ is a subset of $\mathbb{E}(t)$ formed by all the communication links among the vehicles in k. $\mathcal{S}_k(t) = \|\mathbb{V}_k(t)\|$ is the size of the component $C_k(t)$. $\mathbb{C}(t) = \{C_k(t)\}$ refers to the set of components appearing in the network at time t, and $\mathcal{C}(t) = \|\mathbb{C}(t)\|$ denotes their number. We use $C_{max}(t)$ to indicate the largest cluster appearing in the network at time t and $S_{max}(t)$ to represent its size. We denote the number of nodes appearing on the highway at t as $\mathcal{N}(t) = |\mathbb{V}(t)|$. Finally, the average size of clusters appearing at time t is referred to as $S_{avg}(t) = \mathcal{N}(t)/\mathcal{C}(t)$.

In the rest of the paper, for the sake of simplicity, we drop the time notation and we refer to all metrics at a generic time instant: $\mathcal{N}$ represents then the number of vehicles on the highway, $\mathcal{C}$ the number of components, $\mathcal{S}_{avg}$ the average size of component, and $\mathcal{S}_{max}$ the largest component size.



(a) $R = 50$ m (b) $R = 50$ m (c) $R = 100$ m (d) $R = 100$ m

(e) $R = 200$ m (f) $R = 200$ m (g) $R = 300$ m (h) $R = 300$ m

Figure 6: Distributions of the number of components $\mathcal{C}$ (left), and of the largest component size $\mathcal{S}_{max}$ (right), for each trace under different values of R.

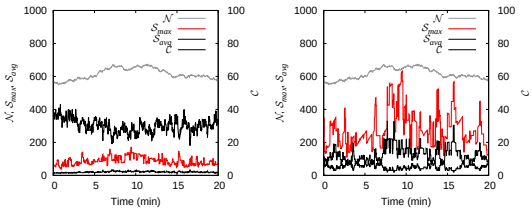
5.2 Instantaneous network connectivity

We start by studying the global connectivity properties of the network at each time instant. Thus, we focus on the distributions of the number of components $\mathcal{C}$ and of the size of the largest component $\mathcal{S}_{max}$. Indeed, $\mathcal{C}$ is a measure of how fractioned the network is, while $\mathcal{S}_{max}$ is the maximum number of nodes that can be reached via multi-hop communication at a given time instant. Clearly, the lower is $\mathcal{C}$ and the larger is $\mathcal{S}_{max}$, the better connected is the vehicular network.

In Fig. 6, we present the distributions of $\mathcal{C}$ (left plots) and $\mathcal{S}_{max}$ (right plots), for various values of the communication range R (on different rows). In every plot, each candlestick refers to one mobility trace, and is obtained by aggregating the metrics computed in all instantaneous graphs observed at every second during the 30-minute timespan of that trace. The lowest and highest values in the candlesticks (in grey) are the minimum and maximum values that the metric attains. The inner error bars (in red) depict the first and ninth deciles, while the box highlights the first and third quartiles (in red) around the median value (in black). Also, the step function in the $\mathcal{S}_{max}$ plots is the maximum value $\mathcal{N}_{max}$ of $\mathcal{N}$, observed in all instances of the respective trace. It thus represents the upper bound to $\mathcal{S}_{max}$: the closer is $\mathcal{S}_{max}$ to $\mathcal{N}_{max}$, the closer the vehicular network is to a fully connected single component.

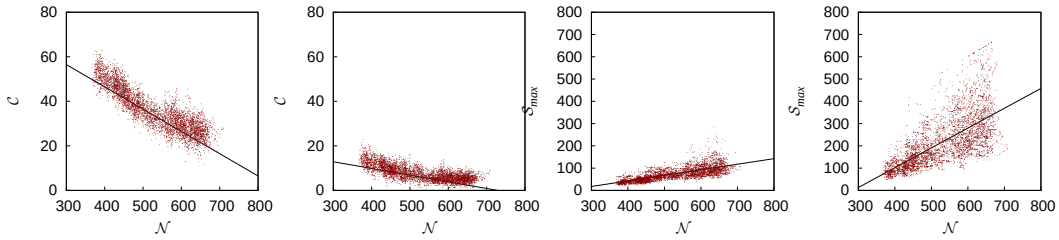
Let us now focus on the case of $R = 50$ m, in Fig. 6(a) and Fig. 6(b). We observe that the connectivity metrics are very similar on different days of the week, forming four clear clusters in both plots. This implies that we can expect the network to have consistent topological properties from Monday to Friday. The aforementioned clusters are instead separated by highway (A6 or M40) and hour (8:30 or 11:30). Specifically, the network appears more fragmented on A6 than on M40, and, in both cases, the connectivity is worse at 11:30 a.m. than at 8:30 a.m. Intuitively, we can ascribe such a variability to the different road traffic conditions encountered on the two highways at different times of the day. In fact, we will further investigate the issue and provide a more rigorous explanation later in this section.

Most of the observations above still hold when considering different communication ranges separately. Nonetheless, when comparing the plots for varying R, it is evident that increasing the communication range can dramatically improve the connectivity of the network. Indeed,



(a) $R = 50$ m (b) $R = 100$ m

Figure 7: Time series of $\mathcal{N}$, $\mathcal{S}_{max}$, $\mathcal{S}_{avg}$, and $\mathcal{C}$ on the A6 highway at 8:30 a.m. of Friday the 7th, for $R = 50$ m (a) and $R = 100$ m (b).



(a) $\mathcal{C}$, $R = 50$ m (b) $\mathcal{C}$, $R = 100$ m (c) $\mathcal{S}_{max}$, $R = 50$ m (d) $\mathcal{S}_{max}$, $R = 100$ m

Figure 8: Scatterplots of $\mathcal{C}$ and $\mathcal{S}_{max}$ versus $\mathcal{N}$, aggregated over all traces.

for $R = 50$ m, there exist on average 30-50 disconnected components, the largest of which only comprises around one tenth of the vehicles. As R grows, however, the network fragmentation is reduced, and more nodes join the largest component. When $R = 300$ m almost all vehicles belong all the time to one single component, as in Fig. 6(g) and Fig. 6(h). We conclude that *the communication range is the foremost parameter controlling the vehicular network connectivity*, as it can induce variations in $\mathcal{C}$ and $\mathcal{S}_{max}$ that are much more significant than those imputable to road traffic conditions. Indeed, $R = 300$ m guarantees a well connected network independently of the traffic scenario, but reducing that value causes the topology to rapidly break apart.

In this latter case, and precisely when R falls below 200 m, differences among the traces emerge. In order to better investigate the causes behind such a diversity, we analyze time series rather than temporal aggregates. More precisely, we observe the evolution of $\mathcal{N}$, $\mathcal{S}_{max}$, $\mathcal{S}_{avg}$ and $\mathcal{C}$ over time, for each trace and varying R. In Fig. 7, we show two sample plots for one of the A6 traces at 8:30 with $R = 50$ m and $R = 100$ m. We omit results for $R = 200$ m and $R = 300$ m because they lead, as seen before, to a stable network with constant behaviors over time (i.e., $\mathcal{C} \sim 1$, and $\mathcal{S}_{max} \sim \mathcal{S}_{avg} \sim \mathcal{N}$). The key remark here is that the time series of $\mathcal{C}$, $\mathcal{S}_{max}$ and $\mathcal{S}_{avg}$ are combinations of two dynamics: slow variations somehow similar to shadowing in a RF signal at the receiver, and rapid variations comparable to fast fading in the RF signal analogy. Next, we analyze these two dynamics in detail.

Slow connectivity dynamics. In Fig. 7, we can glimpse that the slow dynamics of $\mathcal{C}$ and $\mathcal{S}_{max}$ seem correlated to that of $\mathcal{N}$. To explore that possibility, in Fig. 8 we draw scatterplots

Table 3: Correlation coefficients of $\mathcal{C}$ and $\mathcal{S}_{max}$ with respect to $\mathcal{N}$

	$R = 50$ m	$R = 100$ m	$R = 200$ m	$R = 300$ m
$\mathcal{C}$	-0.88	-0.72	-0.28	-0.08
$\mathcal{S}_{max}$	0.72	0.69	0.77	0.98

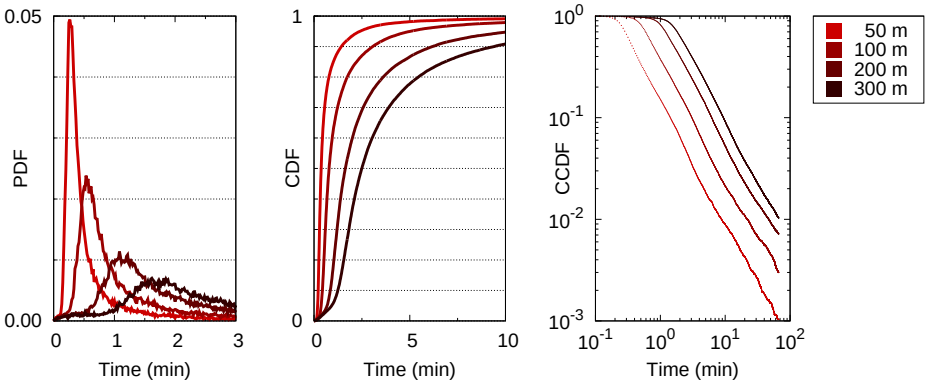


Figure 9: Vehicular contact duration: PDF (left), CDF (middle), and CCDF (right) for M40, on Tue 11 at 11:30 a.m. Curves map to different R values.

of such measures [5] by aggregating data from all of the 16 traces, when $R = 50$ m and $R = 100$ m. There, each point represents one instantaneous connectivity graph in one of the traces. The correlation is evident in all cases. The Pearson product-moment correlation coefficients (PPMCC) [29] computed on the different scatterplots, in Tab. 3, confirm that [6]. Specifically, we remark the strong negative correlation between $\mathcal{C}$ and $\mathcal{N}$ for 50-m and 100-m communication ranges. Such a correlation is reduced when $R = 200$ m and $R = 300$ m, since the number of clusters becomes then very small and nearly constant. For $\mathcal{S}_{max}$, we note a high positive correlation under all communication ranges. We conclude that *the slow topological dynamics of the network are guided solely by the vehicular density*, with more intense road traffic leading to a smaller number of larger components of connected vehicles. This is a key finding, since it unveils how long-term variations of the connectivity are not affected by the type of highway, the weekday or the hour considered, but only by the volume of traffic they yield. It is also interesting to note that this result assimilates highway and metropolitan scenarios, as a akin conclusion was reached for urban vehicular networks [8].

Fast connectivity dynamics The fast dynamics observed in Fig. 7 lead to changes in the connectivity metrics at every few seconds, and are also the cause of the dispersion of values around the linear correlation models in Fig. 8. We speculate that such variations are due to fast-speed vehicular movements that cause links to be established and tore down with high frequency. To investigate this prospect, we compute the distributions of the durations of vehicle-to-vehicle (V2V) contacts in the mobility traces. Due to space limitations, we show the results for one trace only, as all other traces displayed very similar contact duration properties. Fig. 9 portrays the Probability Density Function (PDF), the Cumulative Distribution Function (CDF), and the Complementary CDF (CCDF) of the contact duration between any pair of communicating vehicles over M40, on Mon 11 at 11:30 a.m., with varying R.

As presumed, contacts are relatively short, even if we are considering unidirectional road traffic that should favor long-lived connections [7]. The CDF shows that 80% of the contacts last less than 5 minutes when $R = 300$ m, and less than 1 minute when $R = 50$ m. More importantly to us, there are multiple elements that link V2V contact durations to the fast connectivity dynamics

[5] Note that studying the correlation for $\mathcal{S}_{max} = \mathcal{N}/\mathcal{C}$ would be redundant.

[6] We recall that the value of PPMCC varies between -1 (perfect negative linear correlation) and 1 (perfect positive linear correlation), with values around 0 indicating absence of linear correlation between the variables.

[7] Such long-lived V2V contacts (established by vehicles traveling at similar speeds for a significant amount of time) do exist, but they occur with much lower probability than short-lived V2V contacts. This is demonstrated by the heavy tail of the distributions, shown in the log-log scale CCDF plot of Fig. 9.

we previously observed. First, the probability peak in the PDF for $R = 50$ m is attained at 30 s, i.e., it is in the same order of magnitude of the fast temporal dynamics in Fig. 7(a). Second, the relationship above holds also as R grows, since the fast temporal dynamics become slower in Fig. 7(b) exactly as the V2V contacts become more stable. Third, the effect of higher values of R is milder than one could expect, and the majority of the links last well below one minute at $R = 100$ m. As a result, increasing R generates a much larger number of slightly more stable links, i.e., offers larger space for connectivity disruptions: this is in agreement with the fact that fast connectivity dynamics are stronger in amplitude in Fig. 7(b) than in Fig. 7(a). Overall, these elements let us conclude that *the V2V contact duration is the cause behind the short-term variations of the vehicular network connectivity.*

6 Conclusions

In this paper, we analyzed real-world high-resolution road traffic count measurements, and employed them as the cornerstone for the generation of realistic synthetic traces of vehicular mobility on highways. The resulting datasets are representative of quasi-stationary unidirectional traffic in heterogeneous conditions, including different highways, weekdays and measurement hours. They are the current state of art in publicly available highway traffic datasets for network study.

The topological analysis carried out on the mobility datasets reveals that the factor that primarily controls the connectivity of highway vehicular networks is the communication range. Notably, a 300-m communication range appears to grant full connectivity under any condition; when instead the range falls below 200 m, the network connectivity undergoes temporal variations that are a combination of slow dynamics, controlled by the road traffic density, and fast dynamics, due to the predominant presence of short-lived contacts. Interestingly, we found such properties to be invariant throughout multiple weekdays and on different highways.

References

[1] M. Fiore and J. Härri, "The networking shape of vehicular mobility," *ACM MobiHoc*, 2008.

[2] H.-Y. Huang, P.-E. Luo, M. Li, D. Li, X. Li, W. Shu, and M.-Y. Wu, "Performance evaluation of suvnet with real-time traffic data," *IEEE VTC*, 2007.

[3] M. Doering, T. Pögel, W.-B. Pöttner, and L. Wolf, "A new mobility trace for realistic large-scale simulation of bus-based DTNs," *CHANTS*, 2010.

[4] J. Yuan, Y. Zheng, X. Xie, and G. Sun, "Driving with knowledge from the physical world," *ACM KDD*, 2011.

[5] D. Krajzewicz, G. Hertkorn, C. Rössel, and P. Wagner, "SUMO (Simulation of Urban MObility)," *MESM*, 2002.

[6] J. Härri, M. Fiore, F. Filali, and C. Bonnet, "Vehicular mobility simulation with VanetMobiSim," *Simulation*, 87(4):275–300, 2011.

[7] B. Raney, N. Cetin, A. Vollmy, M. Vrtic, K. Axhausen, and K. Nagel, "An agent-based microsimulation model of swiss travel: first results," *Networks and Spatial Economics*, 3:23–41, 2003.

[8] S. Uppoor, O. Trullols-Cruces, M. Fiore, and J. Barcelo-Ordinas, "Generation and analysis of a large-scale urban vehicular mobility dataset," *IEEE Transactions on Mobile Computing*, to appear.

[9] Y. Pigné, G. Danoy, and P. Bouvry, "A vehicular mobility model based on real traffic counting data," *Nets4Cars/Nets4Trains*, 2011.

[10] S. Uppoor, M. Fiore, J. Härri, "Synthetic mobility traces for vehicular networking," *Vehicular Networks: Models and Algorithms*, H. Labiod, A.-L. Beylot (Editors), Wiley, 2013.

[11] S. Joerer, C. Sommer, and F. Dressler, "Toward reproducibility and comparability of IVC simulation studies: a literature survey," *IEEE Communications Magazine*, 50(10):82–88, 2012.

[12] F. Bai and B. Krishnamachari, "Spatio-temporal variations of vehicle traffic in VANETs: facts and implications," *ACM VANET*, 2009.

[13] M. Caliskan, M. Mauve, B. Rech, A. Luebke, "Collection of Dedicated Information in Vehicular Ad Hoc Networks," *IEEE ITSC*, 2005.

[14] N. Akhtar, O. Ozkasap, and S. C. Ergen, "VANET topology characteristics under realistic mobility and channel models," *IEEE WCNC*, 2013.

[15] D. Naboulsi and M. Fiore, "On the instantaneous topology of a large-scale urban vehicular network: the Cologne case," *ACM MobiHoc*, 2013.

[16] G. Thakur, P. Hui, and A. Helmy, "Modeling and characterization of urban vehicular mobility using web cameras," *NetSciCom*, 2012.

[17] W. Viriyasitavat, F. Bai, and O. Tonguz, "Dynamics of network connectivity in urban vehicular networks," *IEEE Journal on Selected Areas in Communications*, 29(3):515–533, 2011.

[18] S. Yousefi, E. Altman, R. El-Azouzi, and M. Fathy, "Analytical model for connectivity in vehicular ad hoc networks," *IEEE Transactions on Vehicular Technology*, 57(6):3341–3356, 2008.

[19] G.H. Mohimani, F. Ashtiani, A. Javanmard, and M. Hamdi, "Mobility Modeling, Spatial Traffic Distribution, and Probability of Connectivity for Sparse and Dense Vehicular Ad Hoc Networks," *IEEE Transactions on Vehicular Technology*, 58(4):1998–2006, 2009.

[20] M. Gramaglia, I. Soto, C. J. Bernardos, and M. Calderon, "Overhearing-Assisted Optimization of Address Autoconfiguration in Position-Aware VANETs," *IEEE Transactions on Vehicular Technology*, 60(7):3332–3349, 2011.

[21] M. Artimy, W. Robertson, and W. J. Phillips, "Connectivity in inter-vehicle ad hoc networks," *IEEE CCECE*, 2004.

[22] H. Füssler, M. Torrent-Moreno, M. Transier, R. Krüger, H. Hartenstein, and W. Effelsberg, "Studying vehicle movements on highways and their impact on ad-hoc connectivity," *ACM MC2R*, 10(4):26–27, 2006.

[23] M. Gramaglia, P. Serrano, J. Hernandez, M. Calderon, and C. J. Bernardos, "New insights from the analysis of free flow vehicular traffic in highways," *IEEE WoWMoM*, 2011.

[24] B. Kerner, *The physics of traffic*, Springer Verlag, 2004.

[25] M. Treiber, A. Hennecke, and D. Helbing, "Congested traffic states in empirical observations and microscopic simulations," *Phys. Rev. E*, 62(2):1805–1824, 2000.

[26] M. Treiber and D. Helbing, "Realistische mikrosimulation von strassenverkehr mit einem einfachen modell," *ASIM*, 2002.

[27] F. Martelli, M. E. Renda, G. Resta, and P. Santi, "A measurement-based study of beaconing performance in IEEE 802.11p vehicular networks," *IEEE INFOCOM*, 2012.

[28] F. Bai, D. D. Stancil, and H. Krishnan, "Toward understanding characteristics of dedicated short range communications (DSRC) from a perspective of vehicular network engineers," *ACM MobiCom*, 2010.

[29] P. Karl, "Notes on the history of correlation," *Biometrika*, 13, 1920.

Contents

1 Introduction **3**

2 Related Work **4**

3 Real-world traffic count data **5**
3.1 Collecting high-resolution traffic counts 6
3.2 Understanding the data 6

4 Road traffic dataset generation **8**
4.1 Microscopic vehicular mobility models 8
4.2 Parameter calibration 9
4.3 Final dataset 10

5 Connectivity analysis **11**
5.1 Network model and metrics 11
5.2 Instantaneous network connectivity 13

6 Conclusions **16**

RESEARCH CENTRE
GRENOBLE – RHÔNE-ALPES

Inovallée
655 avenue de l'Europe Montbonnot
38334 Saint Ismier Cedex

Publisher
Inria
Domaine de Voluceau - Rocquencourt
BP 105 - 78153 Le Chesnay Cedex
inria.fr

ISSN 0249-0803